

Introduction

Le khi-carré est un test non-paramétrique que l'on utilise pour comparer les fréquences d'une variable catégorielle. Par exemple, une enseignante, appelons-la madame Carmen, enseigne dans une école en campagne et elle veut savoir si les enfants de son école préfèrent le chocolat ou les guimauves. Elle tire 300 enfants de son école au hasard et leur pose la question. Admettons que les 300 enfants répondent qu'ils préfèrent le chocolat. Il est évident que les enfants de son école préfèrent le chocolat, et inversement si les 300 enfants répondent la guimauve, elle en conclut que la guimauve est plus populaire chez les enfants de son école. Maintenant, imaginons que 150 enfants répondent qu'il préfèrent le chocolat et que les autres 150 répondent « guimauve ». On peut en déduire que les deux friandises sont aussi populaires l'une que l'autre. Cependant, si 149 répondent « chocolat » et 151 répondent « guimauve », la guimauve est-elle plus populaire ? Dans ce cas-ci, on peut dire que non, la différence n'est pas suffisamment grande pour dire qu'il y ait une différence significative. La question est : à quel moment la différence devient-elle significative ? Carmen a eu 132 qui ont répondu « chocolat » et 168 qui ont répondu « guimauve ». La guimauve est-elle plus populaire ? Le khi-carré l'aiderait à répondre à la question

Ce test permet de faire une comparaison entre les **fréquences observées** et les **fréquences attendues**. Les fréquences observées sont les fréquences effectivement observées; et les fréquences attendues sont les fréquences que nous nous attendons à observer si l'hypothèse nulle est vraie, c'est-à-dire que toutes les options ont la même fréquence. Pour connaître vos fréquences attendues, il faut diviser son nombre d'observations par le nombre de catégories. Par exemple, dans le cas de Carmen, ce serait $300/2=150$.

Formule

$$\chi^2 = \sum \frac{(O - A)^2}{A}$$

Diagramme explicatif de la formule :

- χ^2 : Symbole du khi-carré
- \sum : Symbole de la somme
- O : Fréquence observée
- A : Fréquence attendue

Mise en situation

Reprenons l'exemple de Carmen. Elle a 300 élèves, dont 132 qui disent préférer le chocolat et 168 qui disent préférer la guimauve. La guimauve est-elle significativement plus populaire que le chocolat ?

H: La guimauve est plus populaire que le chocolat

H₀: La guimauve n'est pas plus populaire que le chocolat

Calculs

	Chocolat	Guimauve	Total
Fréquences observées	132	168	300
Fréquences attendues	150	150	300

Étape 1: Calculer le khi-carré

$$\begin{aligned} \chi^2 &= \sum \frac{(O-A)^2}{A} \\ &= \frac{(132 - 150)^2}{150} + \frac{(168 - 150)^2}{150} \\ &= \frac{-182}{150} + \frac{182}{150} \\ &= \frac{324}{150} + \frac{324}{150} \\ &= 2.16 + 2.16 \end{aligned}$$

$$\chi^2 = 4.32$$

Étape 2: Le degré de liberté

Maintenant, on sait que notre khi-carré est de 4.32, mais on fait quoi avec ça ? Tout d'abord, il faut calculer votre degré de liberté. Dans un cas avec un seule variable comme nous (friandise préférée), il s'agit du nombre de catégories - 1.

$dl = \text{Nombre de catégories} - 1$

$dl = 2 - 1$

$dl = 1$

Étape 3: Distribution khi-carré

Pour cette étape, vous avez besoin de votre table de distribution du khi-carré. Normalement, elle se trouve à la fin de votre manuel ou elle vous a été donnée par votre professeur. Vous voyez la table de distribution du khi-carré fournie par Tabachnick et Fidell dans leur manuel (Tabachnick et Fidell, 2021).

952 APPENDIX C

TABLE C.4 Critical Values of Chi Square (χ^2)

df	0.250	0.100	0.050	0.025	0.010	0.005	0.001
1	1.32330	2.70554	3.84146	5.02389	6.63490	7.87944	10.828
2	2.77259	4.60517	5.99147	7.37776	9.21034	10.5966	13.816
3	4.10835	6.25139	7.81473	9.34840	11.3449	12.8381	16.266
4	5.38527	7.77944	9.48773	11.1433	13.2767	14.8602	18.467
5	6.62568	9.23635	11.0705	12.8325	15.0863	16.7496	20.515
6	7.84080	10.6446	12.5916	14.4494	16.8119	18.5476	22.458
7	9.03715	12.0170	14.0671	16.0128	18.4753	20.2777	24.322
8	10.2188	13.3616	15.5073	17.5346	20.0902	21.9550	26.125
9	11.3887	14.6837	16.9190	19.0228	21.6660	23.5893	27.877
10	12.5489	15.9871	18.3070	20.4831	23.2093	25.1882	29.588
11	13.7007	17.2750	19.6751	21.9200	24.7250	26.7569	31.264
12	14.8454	18.5494	21.0261	23.3367	26.2170	28.2995	32.909
13	15.9839	19.8119	22.3621	24.7356	27.6883	29.8194	34.528
14	17.1770	21.0642	23.6848	26.1190	29.1413	31.3193	36.123
15	18.2451	22.3072	24.9958	27.4884	30.5779	32.8013	37.697
16	19.3688	23.5418	26.2962	28.8454	31.9999	34.2672	39.252
17	20.4887	24.7690	27.5871	30.1910	33.4087	35.7185	40.790
18	21.6049	25.9894	28.8693	31.5264	34.8053	37.1564	42.312
19	22.7178	27.2036	30.1435	32.8523	36.1908	38.5822	43.820
20	23.8277	28.4120	31.4104	34.1696	37.5662	39.9968	45.315
21	24.9348	29.6151	32.6705	35.4789	38.9321	41.4010	46.797
22	26.0393	30.8133	33.9244	36.7807	40.2894	42.7956	48.268
23	27.1413	32.0069	35.1725	38.0757	41.6384	44.1813	49.728
24	28.2412	33.1963	36.4151	39.3641	42.9798	45.5585	51.179
25	29.3389	34.3816	37.6525	40.6465	44.3141	46.9278	52.620
26	30.4345	35.5631	38.8852	41.9232	45.6417	48.2899	54.052
27	31.5284	36.7412	40.1133	43.1944	46.9630	49.6449	55.476
28	32.6205	37.9159	41.3372	44.4607	48.2782	50.9933	56.892
29	33.7109	39.0875	42.5569	45.7222	49.5879	52.3356	58.302
30	34.7998	40.2560	43.7729	46.9792	50.8922	53.6720	59.703
40	45.6160	51.8050	65.7585	59.3417	63.6907	66.7659	73.402
50	56.3336	63.1671	67.5048	71.4202	76.1539	79.4900	86.661
60	66.9814	74.3970	79.0819	83.2976	88.3794	91.9517	99.607
70	77.5766	85.5271	90.5312	95.0231	100.425	104.215	112.317
80	88.1303	96.5782	101.879	106.629	112.329	116.321	124.839
90	98.6499	107.565	113.145	118.136	124.116	128.299	137.208
100	109.141	118.498	124.342	129.561	135.807	140.169	149.449

Source: Adapted from Table 8 in *Biometrika Tables for Statisticians*, vol. 1, 3d ed., edited by E. S. Pearson and H. O. Hartley (New York: Cambridge University Press, 1958).

Une fois votre table de distribution du khi-carré en mains, vous avez besoin de votre degré de liberté et de votre seuil de rejet pour trouver votre valeur critique. Habituellement, on travaille avec un seuil de rejet de 0.05. On a déjà calculé le degré de liberté qui est de 1. Avec ces informations, vous pouvez trouver votre valeur critique dans la colonne de votre seuil de rejet et la ligne de votre degré de liberté.

TABLE C.4 Critical Values of Chi Square (χ^2)

df	0.250	0.100	0.050	0.025	0.010	0.005	0.001
1	1.32330	2.70554	3.84146	5.02389	6.63490	7.87944	10.828
2	2.77259	4.60517	5.99147	7.37776	9.21034	10.5966	13.816
3	4.10835	6.25139	7.81473	9.34840	11.3449	12.8381	16.266
4	5.38527	7.77944	9.48773	11.1433	13.2767	14.8602	18.467
5	6.62568	9.23635	11.0705	12.8325	15.0863	16.7496	20.515
6	7.84080	10.6446	12.5916	14.4494	16.8119	18.5476	22.458
7	9.03715	12.0170	14.0671	16.0128	18.4753	20.2777	24.322
8	10.2188	13.3616	15.5073	17.5346	20.0902	21.9550	26.125
9	11.3887	14.6837	16.9190	19.0228	21.6660	23.5893	27.877
10	12.5489	15.9871	18.3070	20.4831	23.2093	25.1882	29.588
11	13.7007	17.2750	19.6751	21.9200	24.7250	26.7569	31.264

Dans ce cas, notre valeur critique est de 3.84. Cela signifie que si votre χ^2 est **supérieur** à 3.84, vous pouvez rejeter votre hypothèse nulle. Si vous rejetez l'hypothèse nulle, cela signifie que la guimauve est plus populaire que le chocolat. Notre χ^2 est de 4.32, on peut rejeter l'hypothèse nulle.

Signification des symboles

X : Valeur

N : Nombre d'observations total

n : Nombre d'observations d'un groupe

- n_1 : nombre d'observations du groupe 1 ;

- n_2 : nombre d'observations du groupe 2 ;
- etc.

K : Nombre de groupes

Σ : Somme

\bar{X} : Moyenne d'un échantillon

μ : Moyenne d'une population

s : Écart-type d'un échantillon

σ : Écart-type d'une population

s^2 : Variance d'un échantillon

σ^2 : Variance d'une population

χ^2 : Khi-carré

A : Fréquence attendue

O : Fréquence observé

L : Ligne

C : Colonne

t : Statistique t (ou score t dans le cas d'une corrélation/régression linéaire)

r : Coefficient de Pearson

Z : Score Z

b : Coefficient de régression

a : Ordonnée à l'origine

\hat{Y} : Valeur de la VD qu'on veut prédire l'aide la VI

F : Statistique F

Ouvrages de référence

- Field, A. (2017). *Discovering Statistics Using IBM SPSS Statistics: North American Edition* (5th ed.). Sage Edge
- Howell, D.C. (2008). *Méthodes statistiques en sciences humaines* (M. Rogier, V. Yzerbyt, & Y. Bestgen, Trans.). (6th ed.). De boeck. (Original work published 2008)
- Tabachnick, B. G., & Fidell, L. S. (2021). *Using Multivariate Statistics* (7th ed.). Pearson